

PHONETIC MOTIVATION FOR DIACHRONIC SOUND CHANGE IN BANTU
LANGUAGES AS EVIDENCED BY VOICELESS PRENASALIZED STOP PERCEPTION
BY NATIVE SOMALI CHIZIGULA SPEAKERS

Haley Boone

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Linguistics.

Chapel Hill
2018

Approved by:

A. Elliott Moreton

Jennifer Smith

Michal Temkin Martinez

© 2018
Haley Boone
ALL RIGHTS RESERVED

ABSTRACT

Haley Boone: Phonetic Motivation for Diachronic Sound Change in Bantu Languages as Evidenced by Voiceless Prenasalized Stop Perception by Native Somali Chizigula Speakers
(Under the direction of A. Elliott Moreton)

Two hypotheses were tested as triggering nasal effacement (leaving an aspiration contrast) in voiceless prenasalized stops in Bantu languages:

Aspiration is more reliably produced than voiceless nasalization.

Voiceless nasalization is harder to hear than aspiration.

Productions from two Somali Chizigula speakers were measured to test the cue reliability of nasalization amplitude versus aspiration duration. Aspiration is a more reliably produced cue, providing better distinction between voiceless stops.

The perception of voiceless nasalization and aspiration by 10 Somali Chizigula participants was tested. Native productions of voiceless prenasalized and plain stops were cross-spliced to contain pre-burst information from one stop type and post-burst from the other. Participants then identified each stimulus as prenasalized or plain.

Nasalized-only stimuli were identified as “prenasalized” significantly less than control prenasalized stimuli, but aspirated-only stimuli did not receive significantly less “prenasalized” responses than prenasalized controls.

Aspiration appears easier to hear, but not more heavily weighted than nasalization.

Dedicated to my grandmothers, who always taught me to love God, love learning and, where I lacked skill, to at least put forth my best effort.

ACKNOWLEDGEMENTS

First, I give credit to my God and my savior, Jesus Christ, without whom I would never have made it this far. I would like to thank my advisor, Dr. A. Elliott Moreton, for his patience and direction, as well as my other committee members, Dr. Michal Temkin Martinez and Dr. Jennifer Smith, for their advice and critiques. I also want to acknowledge Chris Wiesen at the Odum Institute for his assistance with the statistics presented here, and for his countless attempts to explain basic statistics to me. Thank you to my P-siders, who have faithfully sat through multiple presentations of my work and always provided thoughtful advice and resources; I particularly owe gratitude to Will Carter, who worked through the concept of cue weighting with me over and over again. Thanks to Mengqian Wang for collaborating with me on the spectral tilt project which led to this (and which provided much of the information contained herein). Thank you to my translators and consultants, Mwaliko Mberwa and Dadiri Nuro, and my anonymous Somali Chizigula participants. I would also like to acknowledge University of Alberta for furnishing Actuate. Thanks again to Michal Temkin Martinez for providing the facilities of the Mary Ellen Ryder Linguistics Lab at Boise State University, as well as for her help in running several of the participants in this study. Finally, thanks to my family for always supporting me.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	vi
LIST OF FIGURES	viii
Chapter 1: Introduction	1
1. Summary, Background, and Introduction.....	1
1.1. Summary	1
1.2. Sound and Language Background	3
1.2.1. Prenasalized stops.....	3
1.2.2. Segment status	4
1.2.3. Bantu Languages and Prenasalized stops	5
1.2.4. Historical change in Bantu languages	6
1.2.5. Somali Chizigula Background.....	8
1.2.6. Somali Chizigula in Boise	9
1.3. Study Background.....	9
1.3.1. Aerodynamic Studies.....	9
1.3.2. Acoustic Study.....	10
1.3.3. Huffman and Hinnebusch’s Theory	11
1.3.4. Current Proposal	12
Chapter 2: Cue Reliability	15
2. Cue Weighting and Reliability	15
2.1. Introduction.....	15

2.2. Cue Reliability and Sound Change	19
2.3. Cue Reliability in Somali Chizigula	20
Chapter 3: Perception Experiment	31
3. Perception Experiment	31
3.1. Hypothesis and Predictions.....	31
3.1.1. Hypothesis	31
3.1.2. Predictions	31
3.2. Perception Experiment Methods.....	33
3.2.1. Participants	33
3.2.2. Study Synopsis	33
3.2.3. Materials	34
3.2.4. Procedure	40
3.3. Analysis and Results	42
3.3.1. Analysis	42
3.3.2. Stimuli effects.....	46
Chapter 4: Conclusions	50
4. Discussion.....	50
4.1. Perception and Cue weighting Discussion.....	50
4.2. Implications for Sound Change	51
4.3. Further Study	52

LIST OF FIGURES

Figure 1: Prenasalized Stops Aerodynamic Data.....	10
Figure 2: / ^m p ^h apa/ waveform/spectrogram.....	11
Figure 3: Cue Reliability.....	16
Figure 4: Reliability Sample	17
Figure 5: Shape Assortment.....	17
Figure 6: Color Assortment	18
Figure 7: / ^m p ^h alala/ spectral tilt.....	23
Figure 8: /palapaɬa/ spectral tilt.....	24
Figure 9: / ^m p ^h era/ intensity.....	25
Figure 10: /pera/ intensity.....	25
Figure 11: Reliability of VOT vs. Pre-burst Amplitude: Spectral Tilt	26
Figure 12: Reliability of VOT vs. Pre-burst Amplitude: Relative Intensity	27
Figure 13: F0, F1, F2 and F3 Averages	38
Figure 14: same-spliced / ^m p ^h era/.....	39
Figure 15: same-spliced /pera/.....	39
Figure 16: cross-spliced [^m pera].....	39
Figure 17: cross-spliced [p ^h era].....	39
Figure 18: Results of Perception Study	44
Figure 19: VOTs and Resp. per Item	47

Chapter 1: Introduction

1. Summary, Background, and Introduction

1.1. Summary

This study examined the current state of Somali Chizigula voiceless prenasalized stops and historical data from other languages in an attempt to explain nasal effacement (deletion of the nasalization) in Bantu prenasalized stops. This sound change is claimed to have taken place in other Bantu languages (Huffman and Hinnebusch, 1998).

This paper will explore two theories of sound change: 1. that cue reliability leads to cue weighting – a “preference” for reliable cues and possibly the deletion of unreliable cues (following Toscano and McMurray, 2010) – and 2. that the difficulty of perception leads to the deletion of hard-to-hear cues (proposed by Huffman and Hinnebusch, 1998).

Measurements of nasalization amplitude versus aspiration duration in Somali Chizigula voiceless prenasalized and plain stops will be used as evidence for cue reliability leading to sound change.

The results will show that the length of the VOT (aspiration) is a more consistent cue for discriminating between categories than the amplitude of the voiceless nasalization, which does not always rise above the level of background noise. According to Toscano and McMurray (2010), the greater reliability of one cue in identifying contrastive sounds within a language, in theory, results in that cue – in this case, aspiration – being the cue more heavily weighted by speakers. This may eventually lead to the loss of the less reliable cue.

A perception study testing the cues that listeners focus on when discriminating between voiceless plain stops and voiceless prenasalized stops was used to test the hypothesis, proposed by Huffman and Hinnebusch (1998), that devoicing of the nasalization in voiceless prenasalized stops leads to the deletion of the nasalization due to the difficulty listeners experience in hearing the cue. The results of this study show that, when participants have only either aspiration or nasalization to use in identifying voiceless prenasalized stops versus plain stops, stops with only nasalization lead to significantly less “prenasalized” responses than having both cues. Stops which lack nasalization but have significant aspiration (more than 60 ms.), however, do not lead to significantly less “prenasalized” responses than stimuli containing both cues.

Taken individually, both the cue weighting measurements and the perception study predict that Somali Chizigula will follow the same sound change that has taken place in other Bantu languages. Each, therefore, offers a possible explanation for nasal effacement as it has been known to have taken place already.

However, comparing the cue reliability to the perception study results, one would expect to see a greater difference between the nasalized-only and aspirated-only stimuli than one does if aspiration is weighted more heavily than nasalization. It appears that participants will accept either the presence of nasalization or aspiration as an indication that a sound is prenasalized, without taking into account the brevity of the aspiration, as is expected if they are basing their identification of prenasalized and plain stops mainly on aspiration length.

The paper will organize the information as follows:

Chapter 1 introduces prenasalized stops – the sound which is claimed here to have often undergone a change to simple aspirated stops – and discusses the status of the sound in Somali

Chizigula as it compares to other Bantu languages. This chapter will also explain research already done on prenasalized stops in Somali Chizigula, which led to the current study.

Chapter 2 will introduce the role of Cue Reliability in sound change and discuss the reliability of relevant cues in Somali Chizigula voiceless prenasalized stops, in comparison with voiceless plain stops in the language.

Chapter 3 explains the main hypothesis, that removing nasalization from prenasalized stops will not lead to significantly less “prenasalized” responses than true prenasalized stops, but that when aspiration is removed from prenasalized stops, participants will identify the sound as prenasalized significantly less than actual prenasalized stops. This chapter will then outline my predictions about the perception study results, the methods of the study and the actual results.

Finally, Chapter 4 will discuss the results of the perception study and cue reliability in light of the predicted sound change, and will suggest further avenues of study that could build on the information presented here.

1.2. Sound and Language Background

1.2.1. Prenasalized stops

Prenasalized stops are usually analyzed as a single segment, consisting of a period of nasal airflow followed by an oral stop release (Maddieson and Ladefoged, 1996, Maddieson, 2003).

Usually, where found, this nasalization is voiced, regardless of the voicing of the consonant. The articulatory difficulty of producing nasalization prior to a voiceless consonant is pointed out by many, including Huffman and Hinnebusch (1998), Pater (1999), and Myers (2002). This difficulty of articulation, it will be assumed here, along with the perception of the resulting “misarticulation,” has led to the devoicing of the nasal portion in languages which historically

had voiceless prenasalized consonants, even if the voiceless state of nasalization was never captured by researchers in that particular language. The study presented here will argue that this devoicing of the nasalization, due to the difficulty in perceiving it, in turn led to the current inventories of many Bantu languages, which lack voiceless prenasalized consonants but contain their voiced counterparts.

1.2.2. Segment status

Proof of prenasalized segments being, in fact, segments, rather than clusters, is phonological in nature rather than phonetic. That is to say, according to Maddieson and Ladefoged (1993) there is no line that can be drawn between prenasalized stops and NC clusters based on the phonetics of the sound. Length of the sound, which is often cited as a test for segmentalization, they suggest, is not a reliable test, citing that different simple consonants within a single language may vary considerably in length. Voicing assimilation is also not a valid measure, as individual segments often undergo assimilation to adjacent sounds in various languages, with no claims that they are actually realized as part of the adjacent sound.

Thus, according to Maddieson and Ladefoged, the only way to claim segment status for prenasalized stops is on a phonological basis. These sounds have, therefore, been subjected to syllable-splitting tasks, compared with other consonant clusters (or lack thereof) in the language in question, and so on, in hopes of proving the existence or non-existence of this segment. These tasks lead most to accept prenasalized stops as a single segment, at least in its surface form (see Casali, 1993; Downing, 2005 for arguments on segment status, or Tak, 2011 for discussion on underlying forms).

The analysis presented here of nasal effacement being the loss of one cue, contingent on the fact that another “redundant” cue is available to identify this sound, makes more sense if all cues in

question are analyzed as part of a single segment, rather than being the deletion of one segment in a consonant cluster. Therefore, I will continue to refer to prenasalized stops, in this paper, as a single segment, following the tradition laid out by others who have studied this language and similar sounds in related languages (Maddieson and Ladefoged, 1993; Casali, 1995; Hinnebusch, 1975; Temkin Martinez and Rosenbaum, 2017).

1.2.3. Bantu Languages and Prenasalized stops

Prenasalized stops are well-documented in Bantu languages, with detailed acoustic studies of the sound as it is manifested in different languages done by Maddieson and Ladefoged (Maddieson and Ladefoged, 1993; 1996; Maddieson, 2003). The voiced prenasalized stops are particularly prevalent in Bantu languages, with many fewer languages containing a voiceless prenasalized stop. Those which do not have voiceless prenasalized stops often have aspirated stops, particularly word- and stem-initially, contrasting with unaspirated stops and/or word-medial prenasalized stops or nasal-consonant clusters (Nurse and Philippson, 2003).

Prenasalized stops in Bantu languages are the result of a morpheme reanalysis in Proto-Bantu. Bantu languages have 18 noun classes, each which takes an obligatory prefix. The noun classes in question, 9 and 10, were historically marked by mu- and ni-, but deletion of the vowels resulted in nasals prefixing to stems beginning with other consonants, most often stops (Schadeberg, 2003). Over time, different Bantu languages have reanalyzed this cluster in different ways, but often as a prenasalized stop. In a few languages, prenasalized fricatives have also been attested, but these are particularly prone to misperception, leading to rapid nasal effacement (Busa and Ohala, 1995; Hinnebusch, 1975).

It has been proposed (Tak, 2011; Temkin Martinez and Rosenbaum, 2017) that prenasalization in Bantu languages is influenced by *NC_o (Pater, 1999), a phonological constraint that disallows

nasal segments preceding voiceless consonants. Bantu languages have a tendency to allow only CV or V syllables, and possibly syllabic nasal syllables. Clusters are generally not allowed in these languages. To satisfy the constraint against nasal-consonant clusters, or, arguably, against clusters in general (see Archangeli, et al., 1998 for arguments about this being the outcome of cluster reduction in general), nasals are either syllabic before consonants in many Bantu languages (Somali Chizigula included), or the nasal and the following consonant are combined into a single segment. This second correction results in prenasalized consonants, usually stops, which are generally produced with a voiced nasal portion, regardless of the voicing specifications of the following consonant.

The current state of prenasalized stops in Somali Chizigula, where, at least word-initially, the nasalization is voiceless (partial voicing is retained word-medially), appears to be the step between pre-nasalization of the stops and deletion of the nasalization (Temkin Martinez and Boone, 2016), which, when analyzed in terms of difficulty of perception and compared with the sound inventories of other languages, suggests a cause for the sound change concluded to have already taken place in many Bantu languages. This phonetic realization, in turn, leads to reanalysis and diachronic sound change (usually deletion of the nasal portion), due to the difficulty of hearing voiceless nasalization and the fact that aspiration provides an accurate secondary cue for distinguishing the phoneme, making the nasalization superfluous.

1.2.4. Historical change in Bantu languages

Perusal of Nurse and Philippson's compilation of information on Bantu languages (2003), with chapters regarding different languages provided by several Bantu linguists, provided the information to follow on Bantu language inventories.

Bantu languages tend to have plain stops in their inventory, sometimes, though not always, with a voicing contrast (i). A subset of these languages also allow voiced prenasalized stops in contrast with plain stops. These prenasalized stops occur in more languages word-medially than word- or stem-initially (ii, iii). In a smaller subset, voiceless prenasalized stops occur word-medially with voiced nasalization. In these languages which have voiced but no voiceless prenasalized stops word-initially, there is often an aspiration contrast word-initially (iii, iv). Some languages have voiceless prenasalized stops with voiced nasal airflow in both positions (v), and a very few show varying degrees of devoicing of the nasal, particularly word- and stem-initially, and post-burst aspiration (vi) (Nurse and Philippson, 2003).

Table 1: Bantu Stop Inventories

Bantu Language Stop Inventories:	Word-initially	Word-Medially
i. Very common	{p, (b)}	{p, (b)}
ii. Rather Common	{p, (b)}	{p, (b), ^m b}
iii. Less Common	{p, b, ^m b, p ^h }	{p, b, ^m b, p ^h }
iv. Somewhat Rare	{p, b, ^m b, p ^h }	{p, b, ^m b, ^m p}
v. Even Rarer	{p, b, ^m b, ^m p}	{p, b, ^m b, ^m p}
vi. Extremely Rare	{p, b, ^m b, ^m p ^h }	{p, b, ^m b, ^m p ^h }

Notably, there are no inventories with both aspirated stops and voiceless prenasalized stops allowed in the same position (to the author's knowledge). Given this trend, it has been proposed by various researchers (Hinnebusch, 1975; Contini-Morava, 1997; Maddieson, 2003) that historically, in many other Bantu languages, following the deletion of the vowel in class 9 and 10 prefixes and subsequent fusion of the nasal and the following consonant, the voiceless counterpart of voiced prenasalized stops has undergone effacement of the nasal portion, leaving an aspiration contrast between words which historically had a nasalization contrast. This is particularly prevalent in word-initial stops. Because of this, it has been proposed (Martinez and Boone, 2016; Temkin Martinez and Rosenbaum, 2017) that the current state of the voiceless,

prenasalized stop as we see in Chizigula is an intermediate step on the way to effacement. The phoneme which historically included a voiced nasal portion before a voiceless stop has at this point undergone full devoicing word-initially, and will eventually undergo deletion of the voiceless nasal. Thus, Somali Chizigula provides the opportunity to test theories about Bantu nasal effacement specifically, and sound change in general – at least well-attested, one-way sound changes.

1.2.5. Somali Chizigula Background

Somali Chizigula (also known as Mushungulu, Mushunguli, or Zigua) is a Bantu language spoken by the descendants of Kizigua-speakers brought to Somalia from Tanzania as slaves some two hundred years past (Temkin Martinez and Rosenbaum, 2017). After escaping, they built a community in the Lower Jubba Valley of Somalia, and since then the two language varieties (Tanzanian Kizigua and Somali Chizigula) have diverged to the point of being mutually unintelligible.

Somali Chizigula has a rather large inventory of stops, including plain stops, nasal stops, implosives and prenasalized stops. There is a voicing contrast in plain and prenasalized stops. A table showing the stop inventory follows:

Table 2: Stop types in Somali Chizigula

Stop types:	Bilabial	Alveolar	Velar
Prenasalized	^ɱ p ^h m̥b	^ɲ tʀ n̥d	^ŋ k ^h ŋg
Plain (plosives)	p b	t d	k g
Nasal	m	n	ŋ
Implosives	ɓ	ɗ	ɠ

Note that the alveolar prenasalized stop is followed by a voiceless alveolar trill, rather than the normal aspiration indicated by superscript [h]. This is the only time that this sound occurs in the

language, and is believed to be a realization of the aspiration of the prenasalized alveolar consonant (Temkin Martinez and Boone, 2016).

1.2.6. Somali Chizigula in Boise

In 2011, Mwaliko Mberwa, Jon P. Dayley and Michal Temkin Martinez began documentation of Somali Chizigula at Boise State University (Temkin Martinez and Rosenbaum, 2017). The initial assumption about the state of prenasalized stops in Somali Chizigula was that, like other Bantu languages, Chizigula voiceless prenasalized stops had undergone effacement of the nasal portion of the segment, leaving an aspiration contrast between voiceless prenasalized stops and plain voiceless stops. Intuitions from the native speaker consultant led to the conclusion that there were, in fact, still voiceless prenasalized stops present in the language, which spurred acoustic and aerodynamic studies of these sounds (Temkin Martinez and Boone, 2016; Temkin Martinez and Rosenbaum, 2017).

1.3. Study Background

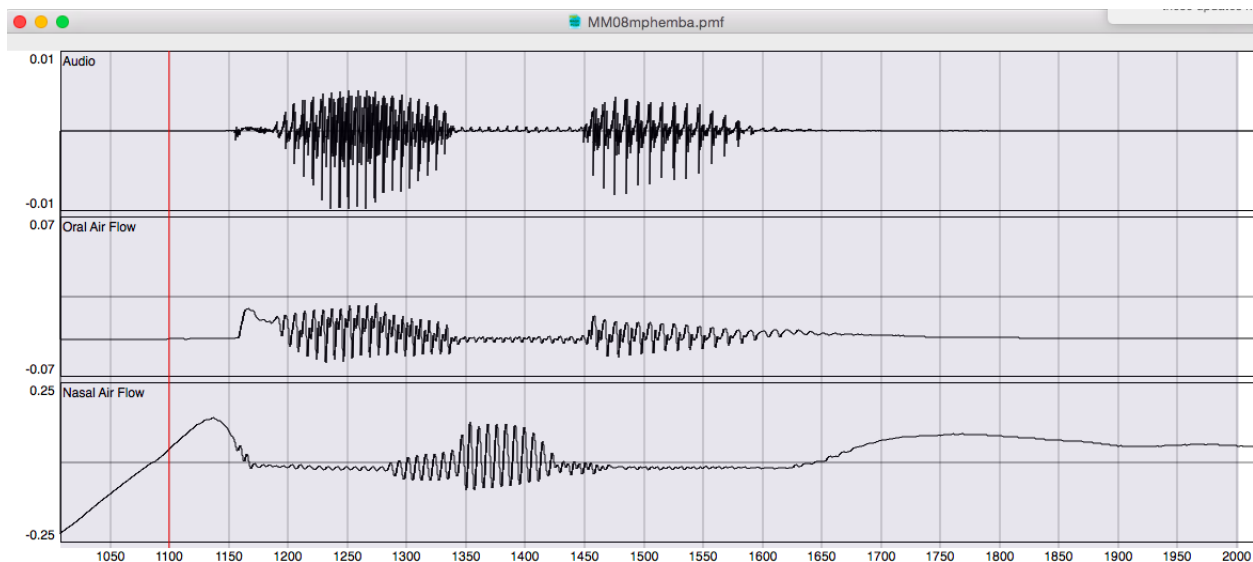
1.3.1. Aerodynamic Studies

Following discovery of the voiceless prenasalized stops in Somali Chizigula, Temkin Martinez led two studies on the sounds, once with Rosenbaum and later with the author (Temkin Martinez and Boone, 2016; Temkin Martinez and Rosenbaum, 2017), which analyzed the acoustic and aerodynamic properties of this sound using nasal and oral masks to measure airflow and voicing during the production of these stops both word-initially and word-medially. Aerodynamic evidence shows robust, voiceless nasal airflow in the production of word-initial voiceless prenasalized stops, lasting between approximately 65 and 100 ms, as well as significant aspiration following the stop, approximately 50-140 ms, whereas the average VOT of plain stops

falls closer to 20 ms. Word-medial voiceless prenasalized stops have partial voicing of the nasal portion. The current study does not focus on the sound in this position, as acoustic cues in this position would differ somewhat from word-initial cues, and voiceless prenasalized stops in this position have proven to be more resilient to effacement than word initially (Contini-Morava, 1997; Nurse and Philippson, 2003)

As is evident in the following figure, nasal airflow (shown in the third channel) spikes at the beginning of the word during the production of [ᵐpʰ], but there is no periodicity, as there is for the word-medial voiced prenasalized stop [ᵐb]. The second channel shows a spike in oral airflow (aspiration) lasting approximately 50 ms. after the release of the stop.

Figure 1: Prenasalized Stops Aerodynamic Data



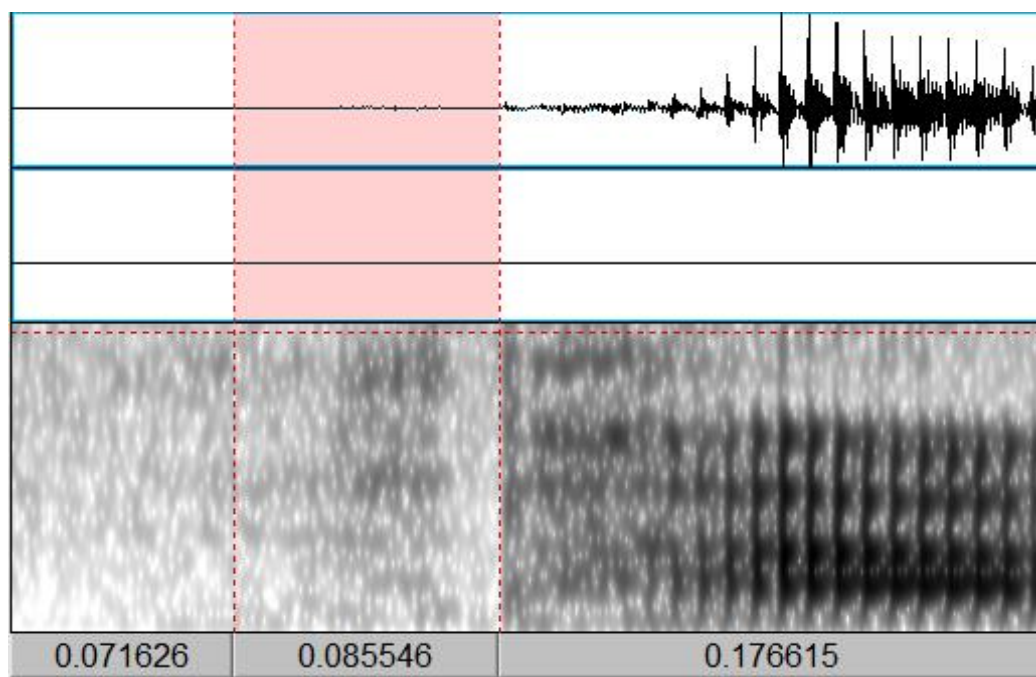
1.3.2. Acoustic Study

Later, an acoustic analysis of the nasal portion of the same segments was done by the author and Wang (2016), who analyzed the spectral tilt and relative intensity of the nasal portion of word-initial prenasalized stops, which showed a low amplitude acoustic signal during production. The

signal during the nasal portion was much lower amplitude than the aspirated portion and is in fact difficult to see in spectrograms and usually not at all evident in waveforms.

The following image, Figure 2: /^mp^hapa/, shows the nasal portion of the word [^mp^hapa] highlighted to show the relative amplitude of the nasal and aspirated portions, evident both by inspection of the waveform, which shows almost no signal during the nasal portion and significantly higher amplitude sound following the release of the stop before the vowel begins, and of the spectrogram during these two portions, which also shows an apparent difference in the amplitude of the two parts, indicated by the darkness of the signal during the aspirated portion.

Figure 2: /^mp^hapa/ waveform/spectrogram



1.3.3. Huffman and Hinnebusch's Theory

Huffman and Hinnebusch (1998) studied voiceless prenasalized stops in Pokomo, a Bantu language spoken in Kenya, and closely related to the language analyzed in this paper. Pokomo voiceless prenasalized stops, similarly to Somali Chizigula, have undergone (partial) devoicing

of the nasal portion of the stop word-initially. Huffman and Hinnebusch state that this is an outcome of imprecision of the timing of gestures involved in the articulation of prenasalized stops, which require the vocal folds to stop vibrating, the velum to raise and the oral cavity to open all in proper sequence (cessation of vocal fold vibration and full velum closure should be simultaneous).

Though Huffman and Hinnebusch do not use this terminology, it seems that they would agree that this leads to the process of phonologization, as defined by Hyman (1972; 2013), whereby natural acoustic consequences of coarticulation become reanalyzed by speakers as part of the phonology – in this case, speakers overextend a devoicing process to the entire nasal portion. Huffman and Hinnebusch's hypothesis is that, when this full devoicing occurs, as they have claimed likely took place in many Bantu languages prior to full nasal effacement, it is often heard as a simple stop, leading to the loss of nasalization altogether.

Thus, Huffman and Hinnebusch claim that sound change (at least the sound change in question) is motivated first by articulatory difficulty, which eventually leads to the prenasalized stop being interpreted as an aspirated stop void of nasalization because of the inability to perceive voiceless nasalization.

1.3.4. Current Proposal

This study will test Huffman and Hinnebusch's hypothesis of nasal devoicing leading to misanalysis of voiceless prenasalized segments to see whether native speakers of a language containing this sound have difficulty hearing the voiceless nasalization, as has been suggested (Huffman and Hinnebusch, 1998).

As mentioned, Somali Chizigula, the language used in this study, has the same voiceless prenasalized stops as are found in Huffman and Hinnebusch's Pokomo, but with totally devoiced nasal air flow word-initially. In addition to containing voiceless nasalization prior to the stop burst, these sounds are produced with significant aspiration following the burst, a cue which is claimed to have developed at the same time as devoicing took place, though which phenomenon caused the other is disputed (see Hinnebusch, 1975; Contini-Morava, 1997; and Hyman, 2003 for discussion of aspiration development in connection with nasal devoicing).

Assuming the difficulty of articulation as the motivation behind devoicing of the nasal portion of voiceless prenasalized stops, and the development of aspiration as a redundant cue, and presuming that the same sound change which has taken place in many related languages began with devoicing in all of those languages, as seen in Somali Chizigula, this study explores the role of perception, and then the possibility of reliability of cue production, as factors leading to reanalysis of these stops as simple aspirated stops – the sound change predicted to follow devoicing of nasalization in these sounds.

It will be claimed in this paper that the sound change from a voiceless prenasalized stop to an aspirated stop, already attested in many languages, is an outcome of the “misproduction” of the nasalization leading to misperception, as suggested by Huffman and Hinnebusch; that related languages have at some point in their history likely undergone a state of voiceless prenasalization similar to the current production of voiceless prenasalized stops in Somali Chizigula; on the basis of three types of evidence: 1. Historical evidence in the form of current sound inventories in related languages compared to reconstructed Bantu forms 2. evidence from the current perception study of aspiration of the stop being a more salient cue than voiceless prenasalization, as well as being the only cue necessary for identification of this sound in contrast to

plain stops in the language, and 3. evidence from the productions of native speakers that cues from aspiration are consistently produced, allowing for identification of each phoneme regardless of whether nasalization is heard.

This study, then, supports the idea that “mishearing” the voiceless nasalization, as we will see in Somali Chizigula, is likely what eventually led to the loss of nasalization in favor of aspiration in Bantu languages.

Chapter 2: Cue Reliability

2. Cue Weighting and Reliability

2.1. Introduction

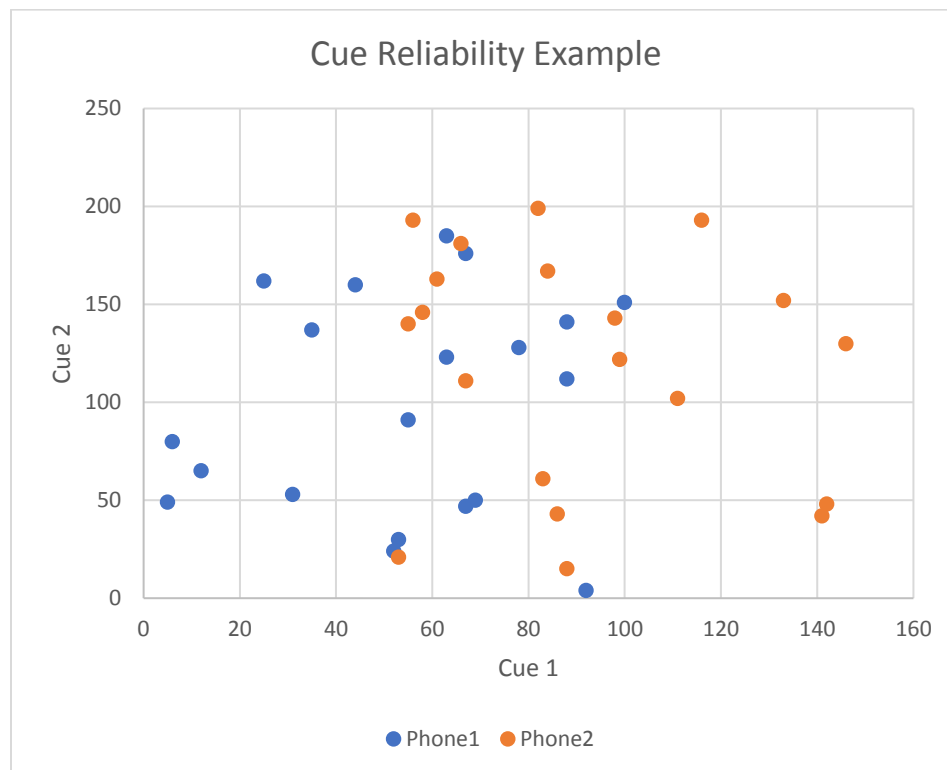
Takagi and Mann (1995), and Cebrian (2006), along with others, have noted the tendency for different speakers (in these cases second-language learners as opposed to first-language speakers) to rely on different cues to identify segments. That is, when there are multiple cues or features which differ between two phonemes, learners can reasonably focus on either cue (or consider both cues equally) when trying to identify which phoneme they are hearing. This has been studied in second language learners of English distinguishing /i/ from /ɪ/, (Cebrian, 2006) where L1 speakers rely more heavily on formants, whereas L2 speakers rely on the duration of the vowel to distinguish these two sounds. Takagi and Mann (1995) found similar results in L1 Japanese learners of English, who used formant cues differently than L1 English speakers in distinguishing /l/ from /r/.

The cue preference that will be explored in this paper involves the perception of nasal airflow versus aspiration in voiceless prenasalized stops, as defined in section 1.2.1.

Toscano and McMurray (2010) investigate cue-weighting as a learned focus on the most reliable cues. They suggest that one of the cues may prove more reliable than others: for example, in the Cebrian study mentioned above, one or the combinations of certain formants may prove more predictable – less variable – in a language. Learners hear these subtle differences and choose

cues with more compact clusters and less overlap between the intended categories to split them; that is, they learn to weight cues based on how well that cue predicts the proper phoneme.

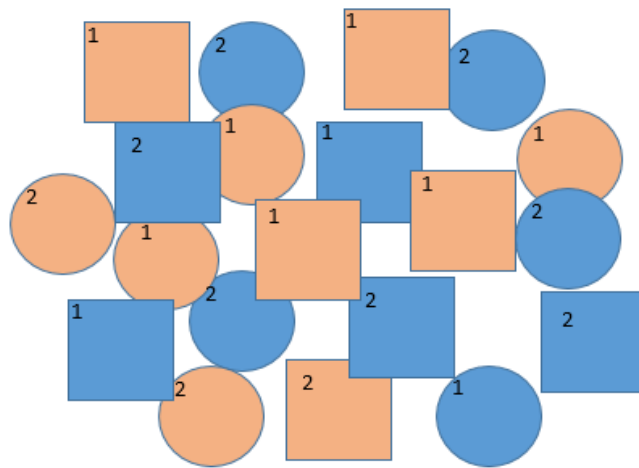
Figure 3: Cue Reliability



Given two sounds: Phone1 and Phone2, which can each be measured in two ways, Cue1 or Cue2, one can categorize these two sounds in two basic ways: by their Cue1 values or their Cue2 values. On the scatterplot above, this would be drawing a line at some value along one axis and assuming that any data point which falls on one side of the line should be identified as Phone1 and data points on the other side of the line as Phone2. On the Cue 1 axis, a border at approximately 75 would rather accurately split Phone1 (indicated by blue dots) and Phone2 (indicated by orange dots). On the Cue 2 axis, on the other hand, there is no point along the axis that would allow accurate identification of the data on one side of the line as Phone1 and data on the other side as Phone2. Therefore, Cue 1 would be a more reliable cue to use in discriminating between these two sounds.

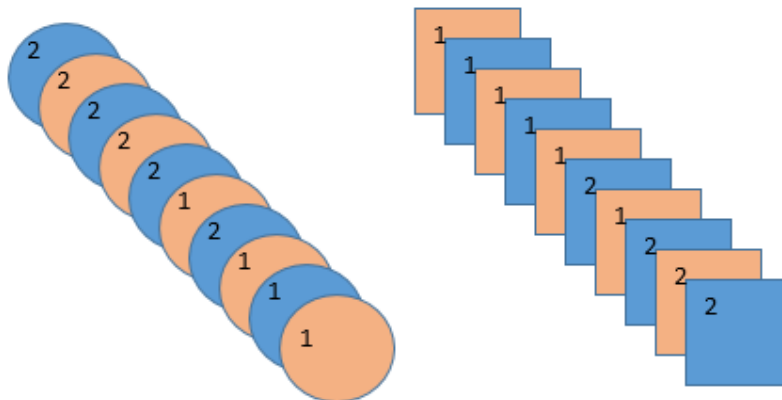
For example, in the following group of items, each which is either a member of set 1 or set 2 (can be thought of as Phone1 and Phone2), as indicated by the number on the top left corner of the item, there are two “cues” that each of these items has: its color and its shape. Not knowing what the number of each item is but knowing there must be two categories, we can organize these shapes into two groups in one of two ways.

Figure 4: Reliability Sample



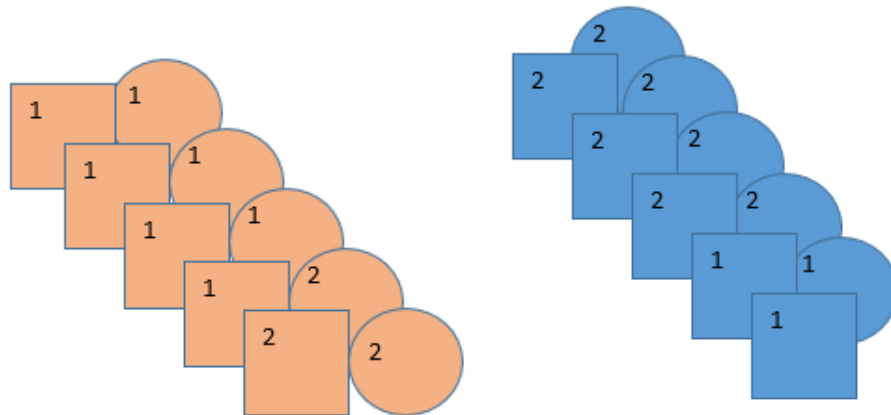
One, we can organize the items by shape. This gives us the following distribution, where each group has six items of one set and four of the other.

Figure 5: Shape Assortment



Or, we can organize them by color, which gives us this distribution, where seven items in each group are from the same set.

Figure 6: Color Assortment



This is a slightly more effective way to split the sets, as it allows us to more accurately group the items into set 1 and set 2. Thus, in terms of cues, color is a better predictor, and thus a more reliable cue, than shape is for identifying members of sets 1 and 2.

This is somewhat simplified, as acoustic cues are not binary. A more accurate example would have some shapes that were somewhere halfway between a square and a circle, and a color that was a mix of orange and blue. The cues we will look at here – VOT (Voice Onset Time) and amplitude – fall at various points along a scale, and there is some overlap between cue values in each category of stop. What we want to know is, essentially, which cue has less overlap between plain and prenasalized stop production, and therefore leads to more accurate categorization of these stops from the listener's point of view. According to Toscano and McMurray's theory (2010), the cue which is more reliably produced should lead listeners to focus more on that cue than a cue that is less reliably produced.

2.2. Cue Reliability and Sound Change

Cue reliability has thus far been discussed as a factor in the weighting of cues by native speakers vs. non-native speakers learning the language, with implications for differences between L1 and L2 perception and phoneme development. Here we will explore the role of cue-reliability and the ensuing weighting of cues as a harbinger of sound change.

Kang (2014) did a corpus study of the VOT and F_0 contrasts between “lenis,” “fortis” and “aspirated” stops in the Seoul dialect of Korean. Korean traditionally has a three-way contrast between voiceless stops: “fortis” [p’], “lenis” [p] and “aspirated” [p^h]. Prior to Kang’s study, there had been considerable research done on the acoustic differences between the categories in this rare, three-way contrast by Han and Weitzman (1970), Cho, Jun and Ladefoged (2002), and others.

Previous studies found that the average VOTs differed between stop types in Korean. According to Cho, Jun, & Ladefoged, the length of VOT is longest for aspirated stops, somewhat shorter in lenis stops, and shortest in fortis stops. However, they also note that lenis stops, which have the highest chance of overlap with both of the other categories in its VOT, as it traditionally has a VOT duration somewhere in between short-lag (fortis) and long-lag (aspirated) stops, were produced with a lower F_0 on the following vowel than the relatively high F_0 of both fortis and aspirated stops, which have the lowest chance of confusability with each other in terms of VOT, as they are at opposite ends of the spectrum.

Kang used this information to do a large-scale apparent time study on the acoustic differences between these stops in the Seoul dialect, comparing the average productions between participants of different genders and ages. According to the results of this study, the duration of VOT is

converging in the lenis and aspirated stops, while the distinction that has been found in the F_0 of the following vowel has become exaggerated in the speech of younger speakers, especially women. Kang concluded from this study that, as VOT duration is becoming less distinct – less reliable in distinguishing between the stop categories – the F_0 is becoming more reliably produced. Thus, the F_0 is becoming more relevant in this distinction, and is expected to become the primary – and eventually perhaps only – cue used in this dialect to distinguish between the lenis and aspirated stops.

This is evidence that sound changes such as that taking place in Seoul Korean, where one cue becomes irrelevant as the other gains relevance, is motivated by, or at least correlates strongly with, the reliability of the production of different cues in the phonemes. Thus, the reliability of nasalization and aspiration was measured in Somali Chizigula voiceless prenasalized stops in comparison to voiceless plain stops, following the prediction that aspiration – the cue which we predict to be preserved – will be more reliably produced than nasalization.

2.3. Cue Reliability in Somali Chizigula

Prenasalized and plain voiceless stops in Somali Chizigula, as mentioned above, differ in two major ways: the length of the aspiration (VOT) and the presence or lack of nasalization prior to the stop burst. Either of these cues seems like it would be sufficient to differentiate between prenasalized or plain stops, and in fact, the results of the perception study discussed later in this paper suggest that speakers can identify the stop types fairly accurately using either of these cues on their own. According to the historical data mentioned in section 1.2.4, Bantu languages tend to choose to keep the aspiration and give up the voiceless nasalization, rather than preserving

voiceless nasalization without aspiration, which leads to the question of why aspiration is so overwhelmingly preferred to nasalization.

The following hypothesis concerning cue reliability in Bantu languages hinges on these claims:

1. Bantu languages have undergone a period of devoicing of the nasalization in prenasalized stops; 2. this devoicing led to the nasalization being misheard – or not heard; and 3. similarly to Seoul Korean, the secondary cue in Bantu (devoiced) voiceless prenasalized stops – aspiration – became a more salient, and more reliably produced cue than nasalization. This section will explore the role of cue weighting in Bantu nasal effacement by measuring and calculating cue reliability and weight in these cues in Somali Chizigula.

Hypothesis

In Somali Chizigula, if the period before the burst (prenasalization vs. plain closure) in voiceless prenasalized and plain stops is measured and compared to the length of aspiration between voiceless prenasalized and voiceless plain stops, aspiration will be found to be the more reliably produced cue.

Then, applying Toscano and McMurray's cue weighting calculations, aspiration should be calculated as the more heavily weighted cue.

Recordings

All cue weighting data comes from the previous experiments mentioned. The acoustic data analyzed in both the Temkin Martinez and Boone experiment and the Wang and Boone study was recorded on the same occasion as the aerodynamic data. Participants were individually asked to say aloud a list of 91 randomized words, repeating each word three times. 32 of these words were relevant to the current study, 16 of which began with a voiceless prenasalized stop, and 16

with a voiceless plain stop, each in varying places of articulation (bilabial, alveolar and velar). Recordings were made in a sound attenuated booth in the Mary Ellen Ryder Linguistics Laboratory at Boise State University using a head-mounted Shure SM-10 microphone and a Zoom H4n recorder. Acoustic data from one male and one female participant recorded in the aerodynamic study (Temkin Martinez and Boone, 2016) was then used in the acoustic analysis done by Wang and Boone (2017). The measurements mentioned in the following sections on nasalization come from the Wang and Boone study. The VOT measurements were made afterward using the same recordings.

Aspiration (VOT)

VOT, or Voice Onset Time, refers to the duration of aperiodic noise (aspiration) between the release of the stop and the following vowel. VOT here was measured in Praat (Boersma and Weenink, 2014) from the end of the apparent burst to the first full cycle of the vowel. The VOTs of each stop type – voiceless prenasalized and voiceless plain – were measured for the recordings taken from the Temkin Martinez and Boone study, and the averages and standard deviations for each was found, shown in the table below. Time here is measured in milliseconds.

Table 3: VOT measurements

VOT (aspiration)	Plain Stops	Prenasalized Stops
Mean (ms)	26	72
Standard Deviation	12.6	19.6
N	33	36

Spectral Tilt

The same sounds were subjected to spectral tilt measurements in a study done by Wang and Boone (2017). Spectral tilt (here referring to change in amplitude of either voiced or voiceless sounds over spectral frequencies) shows essentially how loud a sound is compared to silence.

The spectral tilt is found by looking at a spectral slice and comparing the amplitude at low frequencies, where human speech would occur, to the amplitude at high frequencies, where little to no human speech sounds should be picked up. The greater the difference between the high and low frequencies, the louder the speech sound is. Thus, Figure 7: /^mp^halala/ spectral tilt below, showing a spectral tilt extracted from the nasal portion of the initial prenasalized stop in the word /^mp^halala/, shows a higher amplitude sound than Figure 8: /palapaɬa/ spectral tilt, which shows a spectral tilt extracted from a period during the closure of a word-initial plain stop in the word /palapaɬa/. The horizontal line is set to the same dB level in both images for comparison.

Figure 7: /^mp^halala/ spectral tilt

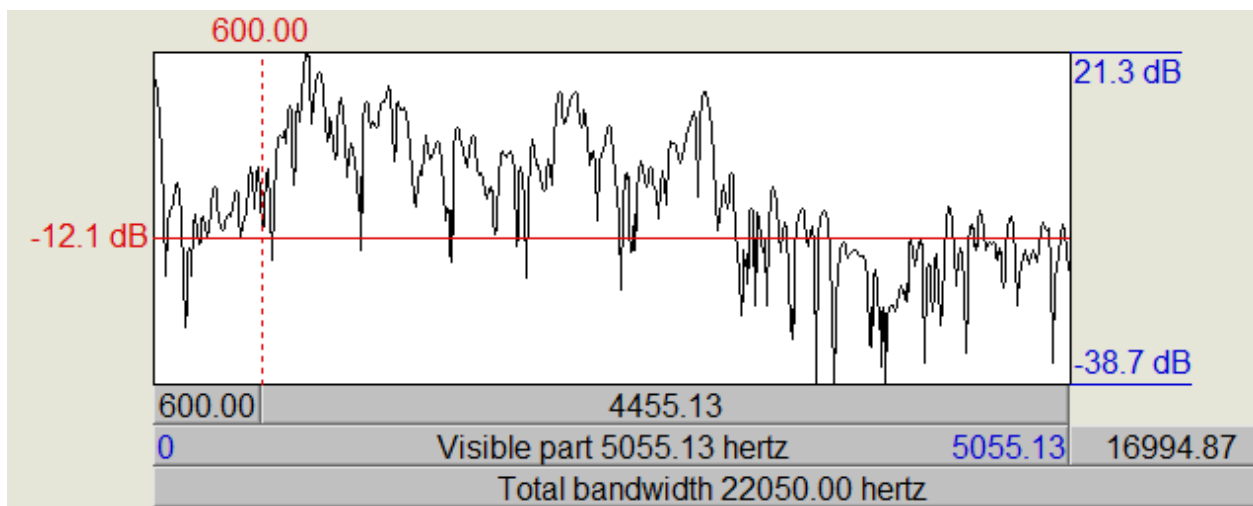
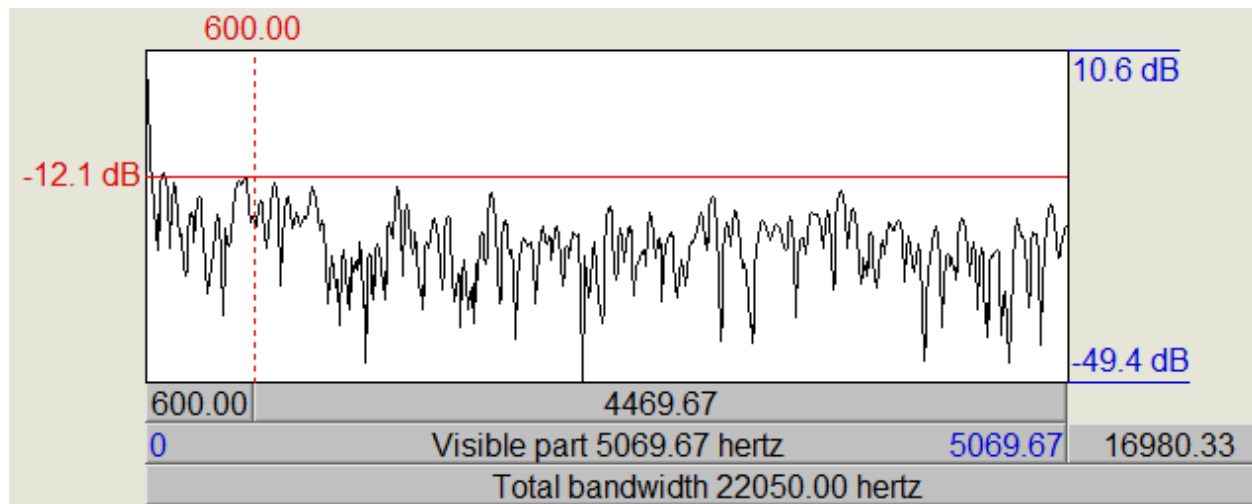


Figure 8: /palapaŋa/ spectral tilt



The difference between the average amplitude between 100 to 600 hz (beginning at the peak just to the right of the leftmost peak and ending at the dotted vertical line) and 4000 to 5000 hz (the rightmost area, about twice the length of the area indicated on the left) was calculated using a function in Praat. Note that, because the lower frequency average was subtracted from the higher frequency average (silence), the tilt averages are negative numbers. Thus, a lower number shows a larger difference, and consequently a louder sound.

Spectral tilt averages are shown in the table below. Measurements for voiced prenasalized stops are shown for comparison.

Table 4: Spectral Tilt Measurements

Pre-burst Amp. (Spec. Tilt)	Voiceless Plain Stops	Voiceless Prenasalized Stops	Voiced Prenasalized Stops
Mean (dB)	-7.5	-13.6	-40
Standard Deviation	4.8	5.1	7.6
N	33	36	44

Relative Intensity

Relative intensity refers to the maximum amplitude of one portion of the word – in this case the nasal portion of word-initial prenasalized stops, or 100 ms. of “silence” before the burst of a

plain stop – compared to another value – here the highest amplitude of the word. This allows the researcher to compare how loud one sound would generally be in relation to another. Wang and Boone (2017) used this method to compare voiced and voiceless nasalization to “silence” in Somali Chizigula. To get these numbers, the difference between the maximum intensity of the whole word (calculated by Praat) and the maximum intensity of the nasal portion or 100 ms. before the release of a plain stop was found, which showed how much quieter each sound was than the stressed vowel. When the numbers for each stop type were compared to each other, it was found that voiceless nasalization was significantly quieter than voiced nasalization, but somewhat louder than silence.

Figure 9: /ᵐp^hera/ intensity

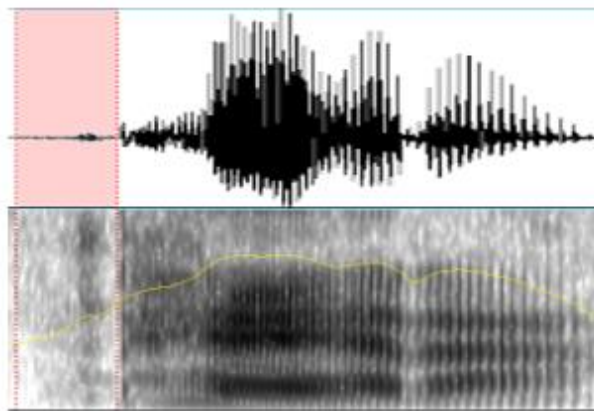
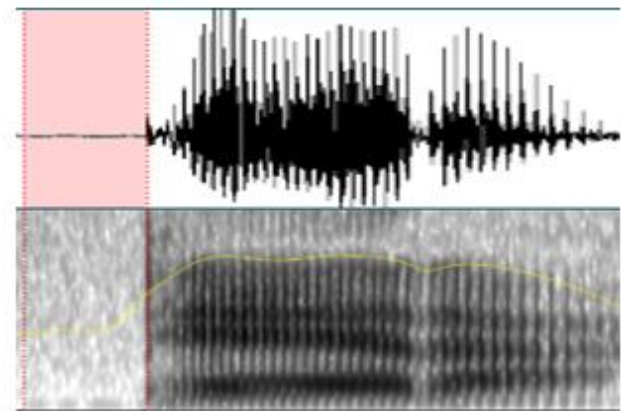


Figure 10: /pera/ intensity



This gave us the results in the following table. Voiced prenasalized stop measurements are included for comparison.

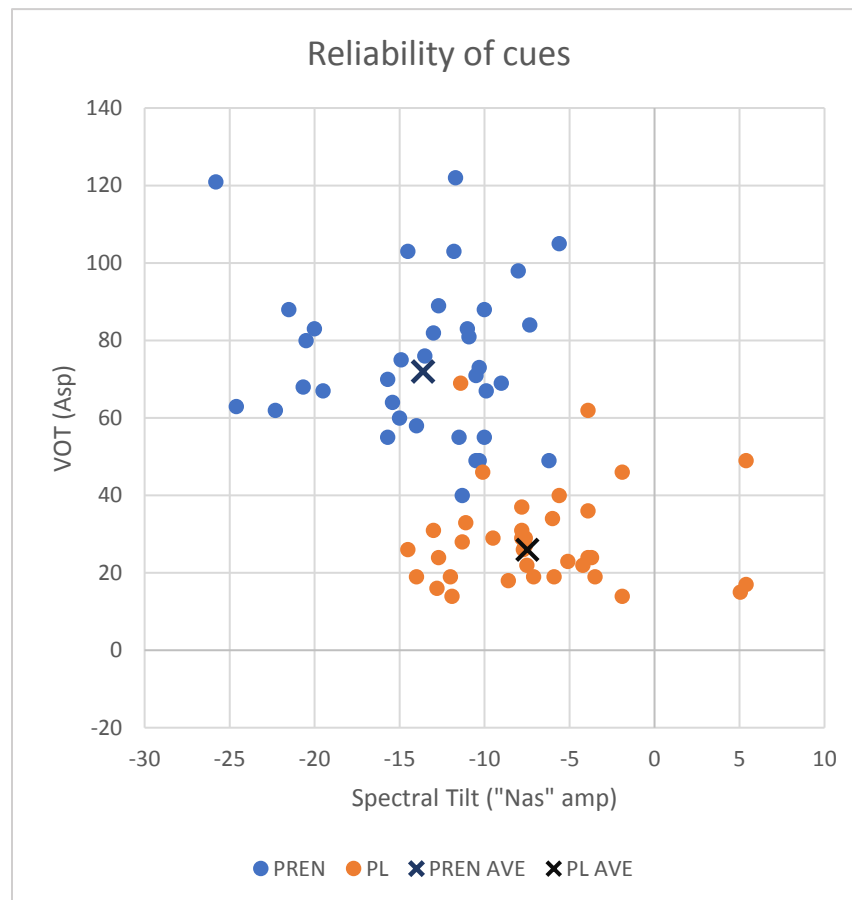
Table 5: Relative Intensity Measurements

Pre-burst Amp. (Rel. Int)	Voiceless Plain Stops	Voiceless Prenasalized Stops	Voiced Prenasalized Stops
Mean (dB)	33.3	29.9	9.0
Standard Deviation	5.8	5.9	4.5
N	33	36	44

Spectral Tilt and Relative Intensity are two different methods used here to measure essentially the same thing – the amplitude of voiceless nasalization (prenasalized stops) compared to silence (plain stops). The spectral tilt values clearly show a more significant difference between the stop types than the relative intensity measurements, but we will consider the results of both methods below.

Figure 11: Reliability of VOT v and Figure 12: Reliability of VOT vs. Pre-burst Amplitude: Relative Intensity (below) show scatterplots of the measurements for stop closure amplitude (intensity) and aspiration length (VOT) compared between plain and prenasalized stops in the language based on spectral tilt or relative intensity data respectively.

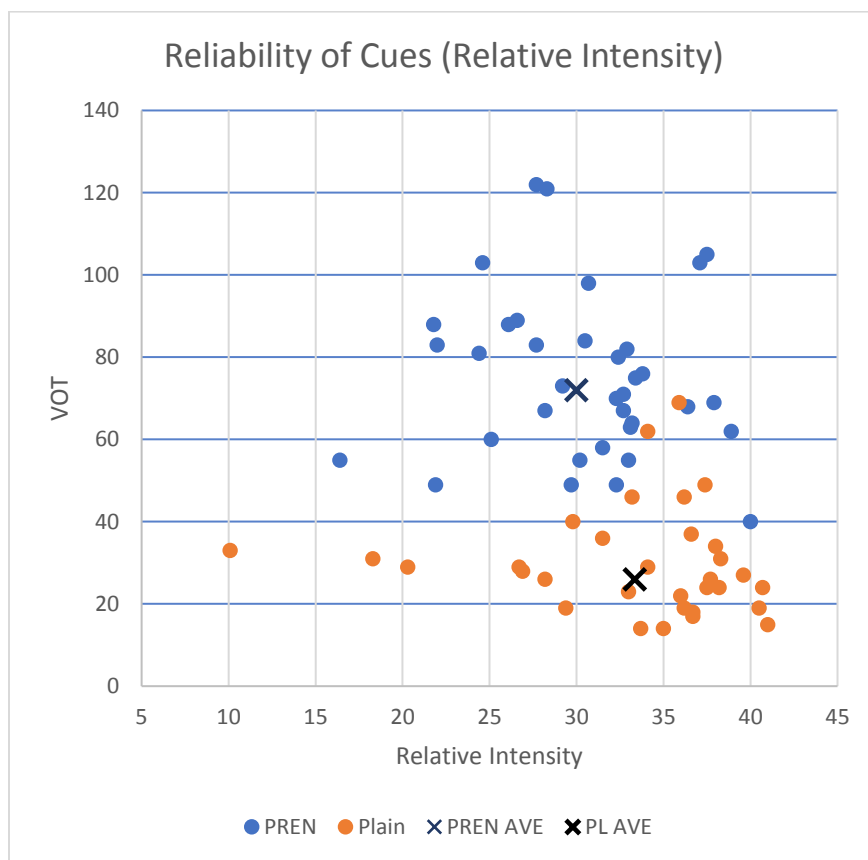
Figure 11: Reliability of VOT vs. Pre-burst Amplitude: Spectral Tilt



From the figure above, we can see that both the spectral tilt of nasalization vs. non-nasalization – aka silence, as occurs before the burst in plain stops – (shown on the horizontal axis) and the VOT of prenasalized vs. plain stops (shown on the vertical axis) show clear clusters for each category of stop (prenasalized stops indicated by blue dots, plain stops by orange). However, there is much more overlap between the two stop types in terms of their spectral tilt than their VOT.

Comparing aspiration to the amplitude of nasalization based on the relative intensity shows less categorization, as shown below, where there is near total overlap between the plain and prenasalized stops in terms of the relative intensity of the pre-burst portion.

Figure 12: Reliability of VOT vs. Pre-burst Amplitude: Relative Intensity



Using either method for measuring nasal amplitude, it is apparent in the above figures that aspiration is a more reliable way to categorize stops in Somali Chizigula than listening for the nasal, which may or may not override background noise (aka “silence”). If the Seoul Korean study done by Kang is any indication, then this should lead to aspiration becoming a more relevant cue than nasalization, which matches the prediction made based on historical evidence that the nasalization will eventually be lost, providing evidence that this was likely what took place in other Bantu languages.

Cue Weighting

Following the hypothesis laid out by Toscano and McMurray that cue-reliability leads to relative weighting of cues, essentially assigning a value to “relevance” as it is discussed by Kang, the values collected above were used to calculate the weight of aspiration as a cue relative to nasal amplitude. The cue-weighting model laid out in Toscano and McMurray (2010),

$$w = (\mu_1 - \mu_2)^2 / \sigma_1 \sigma_2$$

where w is the weight, μ refers to the mean of the category and σ is the standard deviation, was used to assign weights to the two cues – VOT and pre-burst amplitude (based on the more reliable measurements of the spectral tilt). The weight of aspiration as a cue was calculated to be 8.533356, while the weight of amplitude according to the spectral tilt was 1.537152. The weight of the amplitude of nasalization according to the relative intensity measurements was still less, at .337814.

Table 6: Aspiration Weight

VOT (aspiration)	Prenasalized Stops	Plain Stops
Mean (ms)	72	26
Standard Deviation	19.6	12.6
Cue weight	8.533356	

Table 7: Nasal Amplitude Weight (Spectral Tilt)

Pre-burst Amp. (Spec. Tilt)	Prenasalized Stops	Plain Stops
Mean (dB)	-13.6	-7.5
Standard Deviation	5.1	4.8
Cue Weight	1.537152	

Table 8: Nasal Amplitude Weight (Relative Intensity)

Pre-burst Amp. (Rel. Int)	Prenasalized Stops	Plain Stops
Mean (dB)	29.9	33.3
Standard Deviation	5.9	5.8
Cue Weight	.337814	

Following the hypothesis that cue reliability leads to a cue being weighted more heavily than less reliable cues, we should see that Somali Chizigula speakers base their identification of plain vs. prenasalized stops on aspiration without regards (or with less emphasis on) nasalization, as it does not necessarily provide consistent classification of prenasalized stops to the extent that aspiration does.

So far, we have discussed four facts leading us to believe that Somali Chizigula is an example of a midpoint in a common sound change in Bantu languages. 1. Historical evidence in related languages shows that this sound, a voiceless prenasalized stop, has often undergone effacement of the nasalization (particularly word-initially). 2. Aerodynamic data gives evidence that the nasal portion, while still preserved in Somali Chizigula, is totally voiceless word-initially. 3. Acoustic data reveals that voiceless nasalization has rather low amplitude. 4. Upon comparison of voiceless prenasalized and plain stops, there is more overlap in the amplitude of the pre-burst portion than in the length of VOT. It is, thus, reasonable to assume that nasalization would often be lost on listeners in real-world situations, or that aspiration would be weighted more heavily than nasalization, eventually leading to effacement of the nasal portion in favor of the aspiration.

However, the perception study will show, on the contrary, that either nasalization or aspiration on their own provide enough information for native speakers to conclude that a sound is or is not prenasalized. At least at this point, speakers will often identify a stop as prenasalized even when the VOT is short. Thus, although they use a long VOT as evidence of the contrast, a short VOT will not necessarily lead them to classify a sound as a plain stop, as we would expect if they were relying on aspiration and “ignoring” nasalization.

Chapter 3: Perception Experiment

3. Perception Experiment

3.1. Hypothesis and Predictions

Historical evidence of nasal effacement in voiceless prenasalized stops in Bantu languages, as well as the acoustic and aerodynamic data just presented, where nasalization is difficult to see and aspiration is relatively evident, led to the following hypothesis.

3.1.1. Hypothesis

The nasal portion of voiceless prenasalized stops is a less salient cue for identification of the segment than aspiration. Thus, speakers will identify aspirated segments with no nasal portion as prenasalized more often than prenasalized stops where the aspiration has been removed.

I tested this hypothesis by way of a study comparing adult native speakers' perception of the aspiration relative to the nasalization.

3.1.2. Predictions

There are a number of possible outcomes of this study.

1. If aspiration does prove to be a better cue than nasalization, native speakers of Somali Chizigula, when presented with prenasalized stops with only either the aspiration or the nasalization, will identify aspirated stops as prenasalized more often than nasalized, unaspirated stops.

2. If the opposite is true and nasalization is the more apparent cue, then the nasalized, unaspirated stimuli should be chosen as prenasalized more often.
3. It is possible that both cues are necessary, or that either cue serves to indicate pre-nasalization just as well as the other, in which case both types of modified stops should be identified as prenasalized at approximately equal rates, either almost never, if listeners must have both cues, or almost always, if listeners only need one or the other.
4. Finally, participants may simply get confused by the modified stimuli and choose randomly, in which case we will see no particular trend one way or the other.

As will be shown, results of this study show a significant difference between controls and the nasalized-only experimental set, but no significant difference between aspirated-only and the prenasalized controls, and none between aspirated-only and nasalized-only. This seems to support either the first possibility, of aspiration being a more useful cue, or of both cues giving approximately equal chances of identifying the sound (3). A few participants exhibited signs of being in group 4, so their results were removed from the study.

We cannot say definitively, based on these results, that perceptual difficulty is the cause of nasal effacement, which calls for further exploration of possible motivations for sound change, as, one way or the other, something must be motivating the systematic removal of nasalization from this sound.

One such explanation could be production, already noted by Pater (1999), as well as others.

Whether or not the sound is perceptually difficult does not have any bearing on whether it is articulatorily difficult, so the argument could be made, regardless of the results shown here, that phonetics is the motivating factor in the sound change.

3.2. Perception Experiment Methods

As shown above, prenasalized stops in Somali Chizigula differ from plain stops in that they contain both a nasal portion and significant aspiration. This is represented as [^mp^h], in contrast with the plain stop [p]. The purpose of this study was to measure which of these cues (nasalization or aspiration) leads to better identification of prenasalized stops by native Somali-Chizigula speakers.

3.2.1. Participants

Ten native Chizigula-speaking participants took part in this study. Half of the participants were male and half female, all recruited from the Chizigula community in Boise, Idaho. Requirements for participants were that they must speak Somali Chizigula as a first language and have no history of hearing problems. All participants were paid \$10 for their participation in the study, which lasted approximately half an hour per person.

Data from three of the participants was removed from the analysis because they did not meet the proposed number of correct responses for same-spliced tokens (outlined in the Procedure section).

3.2.2. Study Synopsis

For the perception study on voiceless pre-nasalization, using Praat, audio recordings of voiceless plain stops and voiceless prenasalized stops were cross-spliced, swapping the pre-burst portion between a plain and a prenasalized stop (either adding nasalization to a plain stop or replacing nasalization in a prenasalized stop with “silence” from before the plain stop), which produced a sound which is not precisely identifiable as either plain or prenasalized (either [^mp] or [p^h]).

We will hereafter refer to these experimental stimulus types as nasalized-only (or NAS), and aspirated-only (ASP).

Same-spliced tokens – that is, two separate recordings of the same word which have been cut and pasted together without modification of either the aspirated portion or the nasal portion – were also included to be sure that each participant correctly perceives both types of stops in the language ([^mp^h], [p], also abbreviated as PREN and PL, respectively).

3.2.3. Materials

Words and Pictures

12 pairs of minimal or near-minimal pairs (total 24 words, in three places of articulation) beginning with the target sounds were chosen to be used in the study. 10 filler pairs (20 words, in three places of articulation) were also included, all beginning with implosive stops and plain stops. All words were nouns, representable with a picture, and were chosen based on the availability of prenasalized and plain pairs (or plain and implosive pairs for the fillers), with the criteria that each word must be non-violent and deemed likely to be socially acceptable in both American and Somali cultures. Words were selected from the Somali-Chizigula dictionary compiled by Jon Dayley, Mwaliko Mberwa and Michal Temkin Martinez (2016). Target pairs are written below, using the conventions found in the dictionary.

Table 9: Experimental Word pairs

Prenasalized Word	Plain word
mpha “rhinoceros”	pera “pear”
mphalala “corn tassels”	palapacha “perch (fish)”
mphapa “sharks”	papayu “papaya”
mphindi “sections”	pindo “hems”
nkhalala “crabs”	kala “coal/ember”
nkholala “shellfish/snail”	kola “glue”

nkhonde “farms”	konde “slap on the back”
nkhunde “beans”	kundi “bunch/group”
ntrambo “journey”	tambi “branch”
ntrende “date (fruit)”	tende “tent”
ntrongo “sleepy”	tongo “sleep (in the eye)”

Each word was recorded three times spoken by the same native Somali Chizigula speaker in isolation using a head-mounted Shure SM-10 microphone and a Zoom H4n recorder with a sampling rate of 44100 Hz. Recordings were made in the Mary Ellen Ryder Linguistics Lab at Boise State University. The speaker was shown a color picture representing the intended word and an approximate phonetic transcription (based on conventions familiar to the speaker) in point 44 Calibri Light font presented via Powerpoint on an Acer laptop computer with an 8x14 inch screen, and was asked to produce the word in isolation three times consecutively, pausing between each utterance. They then moved on to the next word in the list and did the same for each word.

Pictures representing each word were taken from the internet, based on the English translation for each Somali Chizigula word available in Dayley’s dictionary. A variety of pictures for each word was chosen by the experimenter and then presented to 2 native Somali Chizigula speakers to determine which best represented the intended word.

Pictures were cut into rectangles of the same approximate size, between 2 and 3 inches in length on each side, depending on how much was necessary to keep the full item in the picture, preserving the original background of the photo (items were not cut-out from the background of the picture).

Splicing of Audio

All recordings were normalized in Praat prior to cutting for peak amplitude using .99 as the peak.

The six recordings in each critical word pair (three recordings for each word) were compared for vowel length, and the average calculated. The VOT was compared between the three recordings for each word and the average calculated.

VOTs were measured in Praat using the waveform, measuring from the zero crossing before the first peak in the stop burst to the zero crossing at the beginning of the first clearly visible periodic cycle of the vowel.

Vowels were measured from the zero crossing at the beginning of the first visible periodic cycle to the zero crossing at the end of the last cycle before the consonant in the second syllable began. Because of the varying natures of the following consonant, the exact method differed between stimuli: if a nasal followed, then vowels were measured up to the first antiformant, if a stop followed, the vowel was measured up to the closure, etc.

The two recordings of each word which best fit the average VOT (for that word) and vowel length (for the pair) were used in the same-spliced stimuli, and the most average recording – that is, the recording of each word in the word pair which was closest to the average values for VOT and vowel length, giving length of the vowel precedence, if there was conflict – of each were used in the cross-spliced stimuli. In prenasalized words where the exact boundary between the nasal and the stop burst was unclear, the nasal portion was cut from a recording with a clearer burst.

Two types of stimuli were created from these recordings: 1. cross-spliced, the experimental condition, where part of the recording of one member of a minimal pair was spliced together with a portion from the other member, and 2. same-spliced, which involved splicing together two recordings of the same word as controls.

For the cross-spliced stimuli, the nasal or aspirated portions were swapped with a comparable portion of the other word in the pair (/^mp^hera/ and /pera/ spliced together) Same-spliced stimuli left the category of stops unchanged by replacing a section of a recording with the same part of a different iteration of the same word (i.e. the nasalization in one iteration of /^mp^hera/ was spliced together with the burst of a different recording of /^mp^hera/). Everything beyond the end of the first vowel in the word was cut from the recording for the experimental portion of the study, leaving a single, word-initial syllable.

The boundaries for cutting nasals from prenasalized stops were based on a previous analysis of the aerodynamics of these sounds in Chizigula (Temkin Martinez and Boone, 2016). Each nasal tends to last 100 ms. or less, with an average of approximately 80 ms. For the nasal or non-nasal portion of the word, 100 ms. before the release of the stop was cut (either 100 ms. of silence or containing the nasal) and spliced into the stimuli recording.

The boundary for stops was the beginning of the burst, and everything following the stop in the first syllable (any aspiration as well as the first vowel) was kept. The vowel used in each stimulus, then, comes from the same recording as the burst and, necessarily, the aspiration. This brings up the possible problem of additional cues for pre-nasalization being expressed on the vowel, in the form of length, nasalization, or some other cue.

To test for this possibility, the aerodynamic data in the Temkin Martinez and Boone study (2016) was consulted, which showed no nasal airflow during the production of the vowel. The length of the vowel in each recording was measured, as explained above, and compared between minimal pairs, which also showed no consistent difference.

The F1, F2 and F3 values (reported in Hz) of each vowel was calculated using the Praat formant tracker, with measurements taken from the approximate middle of the vowel, where it appeared stable, and compared to the vowel formants of the other member of the minimal pair, and finally, the F0 of the vowels was also analyzed by measuring the duration of the second full vowel cycle of each recording from zero crossing to zero crossing (reported in Hz). Neither of these measurements showed any consistent differences between the vowels following the different stop types, as shown in the table below, leading to the conclusion, at least for the time being, that we can tentatively assume that speakers are not hearing cues for pre-nasalization in the vowel.

Figure 13: F0, F1, F2 and F3 Averages

Word	Ave F1	Ave F2	Ave F3	Ave F0
mpha	439.3365	1319.274	2319.942	149.8736
pera	453.0197	1405.422	2465.653	126.7230
mphalala	575.7954	981.6381	2196.152	116.4394
pala	623.9858	1029.003	2186.287	122.3587
ntrambo	816.5040	1381.596	2448.079	132.1993
tambi	735.9879	1405.830	2354.602	129.7982
ntrongo	712.7290	1022.546	2352.091	119.2413
tongo	694.0970	1259.667	2332.776	128.3864
nkhalala	758.4158	1329.962	2115.540	120.5056
kala	803.7215	1369.576	2129.949	117.4988
nkholala	610.6977	1000.375	2324.492	123.3773
kola	559.8803	959.6550	2471.865	125.9009

With this in mind, vowels were cut to the average length of both members of the pair (the average of the individual utterances of the word containing the prenasalized stop and the average of the word containing the plain stop), or, in a few cases, lengthened to the average by copying and pasting a cycle of the vowel in the approximate middle of the vowel. Vowels were then tapered using Praat to give a gentler transition into silence. Vowel length was measured from the zero crossing at the first, full, high amplitude cycle of the vowel to the final zero crossing of the last full, high amplitude cycle. The tapering was done using the Praat formula ‘if (xmax-x >

0.015) then self else self * (xmax-x)/0.015 endif”, tapering the last 15 milliseconds of each recording.

Below are Praat images for each type of stimulus. The top pair are same-spliced prenasalized and plain stimuli, respectively. The first image shows both the nasalization and aspiration involved in a prenasalized stop. The second shows a plain stop.

Figure 14: same-spliced /^mp^hera/

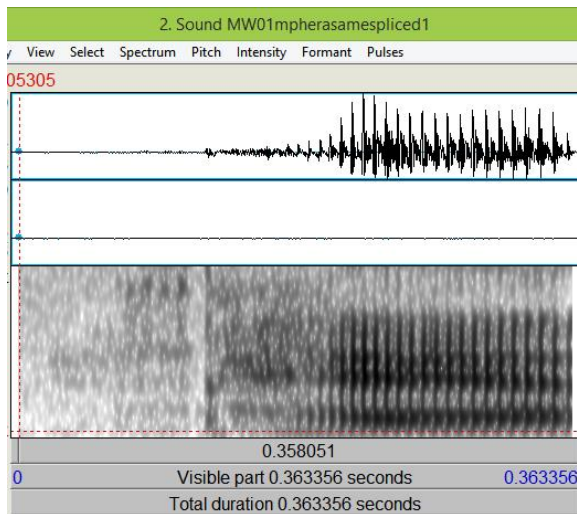
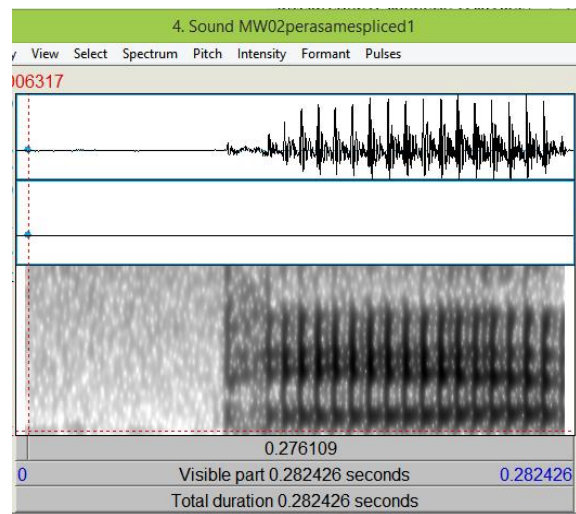


Figure 15: same-spliced /pera/



The second pair of images show cross-spliced stimuli. The leftmost picture shows nasalization but relatively little aspiration. The right picture has significant aspiration but no nasalization.

Figure 16: cross-spliced [^mpera]

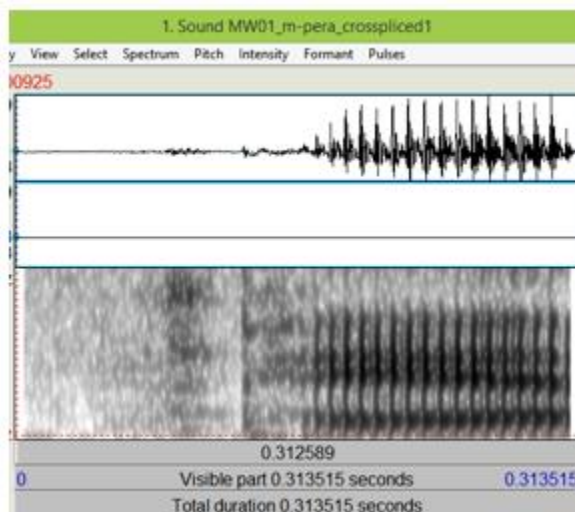
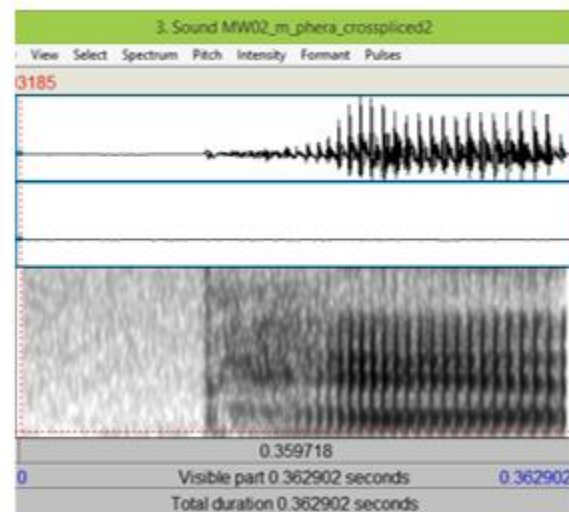


Figure 17: cross-spliced [p^hera]



3.2.4. Procedure

Participants were seated in a sound attenuated booth in front of an Acer laptop computer with a 14x8 inch screen, and were outfitted with sound attenuating headphones and an Olympus LS-10 (Linear PCM) recorder sitting beside the laptop approximately 20 inches from the participant. Both the researcher and an interpreter were present for the duration of the study, and participants were encouraged to ask if they had questions, and to adjust the volume in the headphones to a comfortable level.

The study consisted of three phases: training, practice and an experimental portion.

Phase 1: Training Period

At the beginning of the study participants went through a short training period, where they were shown a color picture via Powerpoint and heard, through headphones, an unmodified audio file containing the intended word as spoken by a native Somali Chizigula speaker. Participants were instructed that they could replay the word as many times as necessary, and to mention if a word was unfamiliar to them (this led to the removal of one pair of words – /^mp^halamunju/ “dragonfly” and /palapatʃa/ “perch (fish)”, where the picture did not match the word).

Participants were then asked to repeat the word that they heard, and their production was recorded using an Olympus LS-10 recorder. This was to ensure that participants knew the word that was intended to be associated with the picture, and that the prenasalized stop is a sound which is produced by the speaker.

Phase 2: Practice

Before the experimental part began, participants practiced identifying a word by listening to only the first syllable and practiced selecting the picture using the indicated keys. The keys “l” and

“d”, were chosen based on their placement on the keyboard, so that the corresponding picture would appear above the key. The keys were marked with a bright pink tag on which the symbols ← and → had been drawn, meant to indicate which side of the screen the corresponding picture appeared on.

They were shown a pair of pictures not used in the experimental portion of the study, but which they had been trained on during the training period, and heard the first syllable of a word which corresponded to one of the pictures. They were asked to press the key the corresponded to the picture that the audio file matched. They went through a series of 12 of these decision tasks, then the section ended and participants were instructed that they would be starting the experimental portion. There was no time limit for participants to decide in either the training or the experimental portion of the study.

Phase 3: Experimental Portion

For the experimental portion of the study participants heard the spliced or unchanged sound ([^mp^h], [p], [^mp] or [p^h]) with a following vowel and saw a pair of pictures presented side-by-side, each representing a word which they had been trained on, one which began with a plain stop and the other which began with a prenasalized stop. The picture representing the prenasalized word always appeared on the right, and the plain stop picture always appeared on the left of the screen. Each pair of words began with the same first syllable (excluding the initial stop) followed by the same consonant in the onset of the following syllable. When possible, they were a minimal pair. Otherwise they were paired by length of vowel and identity of features in the first 3 segments. Participants were asked to select which picture began with the syllable they heard. Each critical pair was cycled through 4 times, once for each audio stimulus.

Ten pairs of words not containing a prenasalized counterpart were recorded and included as filler tokens. Each filler pair was cycled through four times, twice same-spliced and twice cross-spliced. Thus, each participant made 88 decisions for Phase 3 of the study, 48 of them pertinent to the study.

Pictures and audio for the experiment were presented with Actuate software, courtesy of the University of Alberta, which also collected responses and response times for each participant. Pairs were randomly ordered for each participant. The syllable was played twice with a short period (800 ms) of silence between iterations for each decision. A period of 500 milliseconds lapsed between the participant's selection and the presentation of the next pair of pictures, and each audio file played for 100 milliseconds before the burst of the stop (this period contained the nasal portion for the tokens with nasalization).

This study had been run as a pilot study with 2 native Somali Chizigula consultants. Comments from the consultants after the study prompted the normalization of amplitude, repetition of the audio and tapering of vowels.

3.3. Analysis and Results

3.3.1. Analysis

The percentage of times that a prenasalized stop was selected for each of the four types of stimuli of interest (same-spliced prenasalized, same-spliced plain, cross-spliced nasalized-only and cross-spliced aspirated-only) was measured, and the significance of the difference in identification of each type of stimuli was analyzed using a mixed effects logistic regression model to show the probability of answering “Prenasalized” modeled as a function of a single variable of 4 values: Prenasalized (+Nasalization, +Aspiration), Plain (-Nasalization, -

Aspiration), Nasalized only (+Nasalization, -Aspiration) and Aspirated only (-Nasalization, +Aspiration):

$$\text{logit } [p(\text{response} = \text{Pren})] = \alpha (\text{Prenasalized}) + \beta_1(\text{Plain}) + \beta_2(\text{Aspirated}) + \beta_3(\text{Nasalized})$$

Standard errors were adjusted for multiple observations within subjects. Between-subject factors appeared to have no effect. Item effects were not able to be calculated because some items were identified as the same sound by all speakers (giving a 100% response rate).

In the experimental portion of the study, if a participant mislabeled 30% or more of the same-spliced tokens, their data was removed from the analysis. This was the case with three participants, who all produced the prenasalized stop in the training session but appeared to have difficulty hearing it.

The tests yielded the following results, raw percentages shown in Figure 18: Results of Perception Study, statistical test results in Table 10: Statistical Test Results:

Figure 18: Results of Perception Study

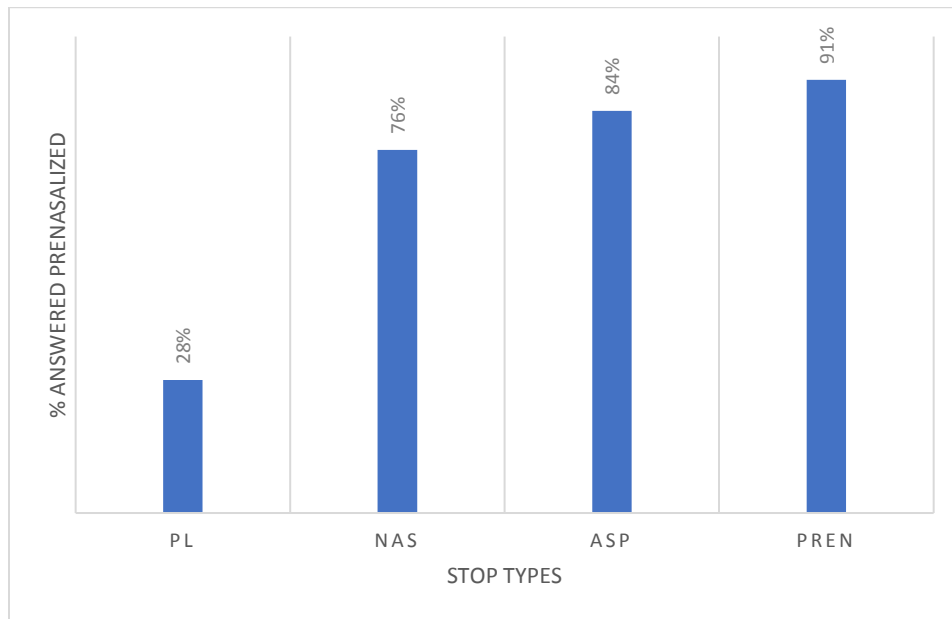


Figure 18: Results of Perception Study, above, shows the percentages of “Prenasalized” answers for each stimuli category, where PL means that the audio presented was a plain stop, PREN were prenasalized stops, and NAS and ASP indicate the experimental cross-spliced stimuli including only the nasal portion or aspirated portion of the prenasalized stop, respectively. All participants and all items are included in the percentage, with the exception of those earlier mentioned as having been removed. Thus, participants in the study identified audio containing a plain stop as being a prenasalized stop 28% of the time, and so on.

Table 10: Statistical Test Results

<i>Label</i>	<i>Mean Estimate</i>	<i>MeanLowerCL</i>	<i>MeanUpperCL</i>	<i>Chi-Square</i>	<i>Pr > ChiSq</i>
ASP	0.8421	0.7874	0.8848		
NAS	0.7632	0.6414	0.8530		
PL	0.2800	0.2002	0.3766		
PREN	0.9079	0.8062	0.9589		
ASP vs. NAS				1.82	0.1770
ASP vs. PL				70.67	<.0001
ASP vs. PREN				1.47	0.2247
NAS vs. PL				28.97	<.0001
NAS vs. PREN				15.42	<.0001
PL vs. PREN				36.08	<.0001

While Figure 18: Results of Perception Study shows a slight tendency for participants to rely on aspiration as a cue for identification of the prenasalized stops rather than the nasalization, the test results above show no significant difference between the experimental groups ASP and NAS ($\chi^2 = 1.82, p = .1770$). There is a significant difference between NAS and PREN ($\chi^2 = 15.42, p = <.0001$), however, whereas the difference between ASP and PREN is insignificant ($\chi^2 = 1.47, p = .2247$), which complicates matters. PL is significantly different from all other categories. This says, essentially, that NAS and ASP have a high probability of being the same category, and ASP and PREN as well, but that NAS and PREN are not the same.

To clarify, in the terms used up to this point, when the nasalization is removed from prenasalized stops, leaving the aspiration (ASP), native speakers do not identify them as prenasalized significantly less than if they have both cues to consider, but when the aspiration is removed, leaving only nasalization (NAS), they are significantly less likely to identify them as prenasalized than if they have both cues. Thus, it appears that removing aspiration causes more

confusion than removing nasalization, but that speakers will accept the presence of either cue as an indication of the prenasalized category (without taking into account the lack of the other cue).

3.3.2. Stimuli effects

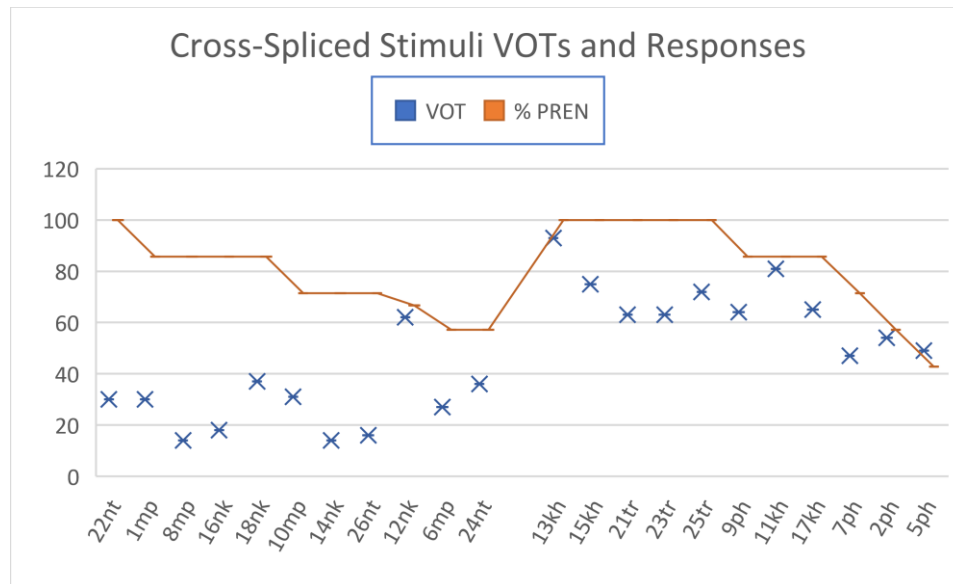
As stated, the between-stimuli effects could not be calculated statistically. Raw results show that a few of the stimuli were identified as prenasalized noticeably more or less than others in its category. For instance, several of the aspirated-only stimuli were identified as prenasalized 100% of the time. One cross-spliced, aspirated-only stimulus resulted in a meager 43% “prenasalized” identification. Its same-spliced, prenasalized counterpart elicited a 57% (correct) “prenasalized” response rate – the lowest rate for the prenasalized audio files (compared to the next lowest at 86%).

Due to this discrepancy, after the study was complete, the individual stimuli were again measured for VOT to evaluate whether the length of VOT correlated with the percentage of “prenasalized” responses. As mentioned in the methods section, the VOT for each stop was untouched when the splicing of the audio files was done, and at times the length of VOT for plain and prenasalized stops showed some overlap. It was found that, in the aspirated stimuli which were most often labeled “plain”, the aspiration always fell around the boundary between prenasalized and plain voiceless stops (50 ms.). The recording used in both the cross-spliced aspirated-only stimulus (ASP) and same-spliced prenasalized stimulus (PREN) which elicited the lowest rate of “prenasalized” responses had a VOT of 50 ms.

The figure below shows the VOT of each experimental stimulus (each labeled with a blue x), as well as the “prenasalized” response rates (represented by the orange line). The set on the left are cross-spliced nasalized-only stimuli, and the right are aspirated-only. Each set is organized in

descending order of response rates, with the stimulus with the highest response rate on the left, and the lowest on the right.

Figure 19: VOTs and Resp. per Item



The cross-spliced nasalized-only stimuli set (on the left in the figure above) had only one stimulus which fell around this point, “12nk” (all other stimuli had VOTs under 40 ms.). This stimulus had a VOT of 62 ms., and was identified by speakers as prenasalized 66% of the time, less than the average for nasalized stimuli, even though it had the longest VOT of the group. The corresponding stimulus in the same-spliced plain category (where the aspiration from both came from the same recording), was incorrectly identified as a prenasalized stop 71% of the time, compared to the next highest of its kind at 43%. Thus, the length of the VOT appears to have somewhat affected response rates for these stimuli.

In the same-spliced prenasalized stimulus which received the lowest “prenasalized” responses, it would seem that the nasalization did not overcome the brevity of the VOT to make consistent identification as “prenasalized” possible. The length of the aspiration necessary for consistent

identification as being “prenasalized” seems to be around 60 ms., although the exact cut-off is uncertain, and may differ according to place of articulation.

In the aspirated stimuli section in the figure above (the right set), the highest “prenasalized” response rates correspond with velar and alveolar stops, with the three lowest response rates falling on bilabial stops. Bilabials tend (cross-linguistically as well as in this stimulus set) to have the shortest VOTs and velars tend to have the longest, so it is difficult to say whether this is an effect of VOT length or some other cue related to place of articulation.

Recall that alveolars are produced with a short, voiceless trill or tap following the burst. This should provide additional, very salient cues for the prenasalized alveolar stops, and, indeed, stimuli which contained the trill elicited a 100% response rate. Interestingly, however, as is apparent in the following table, when the trill was removed and replaced with the plain stop burst and aspiration, in one of the three stimuli the same “prenasalized” response rate of 100% was observed, the highest response rate for a nasalized-only stimulus, showing that the lack of the trill does not prevent listeners from labeling it a prenasalized stop.

Table 11 shows the responses broken down by participant and by token. The leftmost column lists a short code for each cross-spliced token, noting whether the stimuli contained the nasalization (in the upper section, noted by the stop preceded by a homorganic nasal), or aspiration (in the lower section, labeled with “h” following each stop).

Participant codes (in the top row) note the sex of the speaker (male or female) and the order that they were tested in. Boxes filled with black indicate that the participant responded “prenasalized” for that token by that participant, while white indicates a “plain” response. The VOT for each stimulus is noted in the rightmost column.

Total percentages of “prenasalized” responses for each participant are noted in the corresponding column, and the percentage of “prenasalized” responses for each token in the %Pren column.

Table 11: Participant Responses

	1F	2F	3M	4M	5M	6M	7F	%Pren	VOT
NAS	90%	91%	82%	91%	64%	64%	55%		
1mp								86%	30 ms
6mp								57%	27 ms
8mp								86%	14 ms
10mp								71%	31 ms
12nk								67%	62 ms
14nk								71%	14 ms
16nk								86%	18 ms
18nk								86%	37 ms
22nt								100%	30 ms
24nt								57%	36 ms
26nt								71%	16 ms
ASP	80%	73%	91%	91%	82%	91%	82%		
2ph								57%	54 ms
5ph								43%	49 ms
7ph								71%	47 ms
9ph								86%	64 ms
11kh								86%	81 ms
13kh								100%	93 ms
15kh								100%	75 ms
17kh								86%	65 ms
21tr								100%	63 ms
23tr								100%	63 ms
25tr								100%	72 ms
Total	85%	82%	86%	91%	73%	77%	68%		

Black =
prenasalized
response

White = plain
response

NAS =
nasalized-only
cross-spliced
stimuli

ASP =
aspirated-only
stimuli

Chapter 4: Conclusions

4. Discussion

4.1. Perception and Cue weighting Discussion

When the reliability of the production of nasalization versus aspiration in Somali Chizigula stops is calculated, we see that aspiration is a significantly more consistent cue than nasalization in identifying voiceless prenasalized versus plain stops, due to the amplitude of nasalization often failing to rise above background noise. Thus, in the natural world, nasalization may often be lost on hearers, resulting in the eventual loss of this sound, leaving the more consistently produced aspiration as the main cue.

Perception data from this study shows that, in the current state of word-initial, Somali Chizigula voiceless prenasalized stops, having only either aspiration (of more than 50 ms., importantly) or nasalization results in a rather high probability of identification of the sound as a prenasalized stop – at least with good quality recordings presented through headphones in a sound-attenuated room – but that removing aspiration leads to significantly less “prenasalized” responses than having both cues, whereas removing nasalization does not. When nasalization is heard by native speakers of Somali Chizigula (76% of the time), they interpret the sound as being prenasalized regardless of how short the VOT is.

When the results of the perception study and the reliability measurements are combined, it is clear that the reliability of cues has not resulted in aspiration being considered a more important cue than nasalization, as speakers identify nasalized stops with little aspiration as prenasalized

stops, as well. Thus, cue weighting, in this case, has not taken place. It appears that listeners will take either aspiration or nasalization as a cue in identifying prenasalized stops, and that, while one cue is easier to hear, they listen just as hard for the other cue.

Considering the difficulty of the production of a nasal followed by a voiceless consonant in terms of the inability to precisely time articulator movement, as described by Pater (1999), along with the superfluity of having both the nasalization and the aspiration, then simplifying the consonant seems like a phonetically efficient solution. Considering that aspiration is consistently produced and easier to hear than voiceless nasalization, aspiration can be concluded to be the more logical of the two cues to be preserved, although as yet speakers show no partiality toward aspiration over nasalization.

4.2. Implications for Sound Change

The present study looked only at word-initial, voiceless prenasalized stops, which, based on Somali Chizigula and Pokomo, was likely the state of prenasalized stops immediately prior to full nasal effacement in other Bantu languages, leading to an aspiration contrast between what are now word-initial prenasalized and plain voiceless stops in Somali Chizigula. This study has nothing to say, then, about either voiced prenasalized stops or word-medial voiceless prenasalized stops.

In fact, according to the previous studies mentioned (Temkin Martinez and Boone, 2016; Temkin Martinez and Rosenbaum, 2017), at least partial voicing has been preserved in word-medial voiceless prenasalized stops, and full voicing is present in voiced prenasalized stops both word-initially and word-medially (recall that consonants are not allowed word-finally). Because voiced nasalization does not have the same lack of acoustic cues as voiceless nasalization, voiced

prenasalized stops and word-medial voiceless prenasalized stops are not as likely to undergo effacement, or at least not as rapidly. This trend is attested in other Bantu languages, such as some dialects of Swahili, where word-initial voiceless aspirated consonants appear, but word-medially, the nasal in prenasalized stops have been preserved, often through making the nasal syllabic, in spite of its having undergone total nasal effacement word-initially, (Contini-Morava, 1997).

The current state of word-initial voiceless prenasalized stops in Somali-Chizigula is here proposed to be in the semi-final stages of the nasal effacement process, one step further than Pokomo, which has undergone significant devoicing but retains partial voicing, and one step behind Swahili's full nasal effacement.

Following is the observed sound change trend in Bantu languages:

$$*NC \rightarrow NC \rightarrow (NNC \rightarrow NNC^h) \text{ or } (NC^h \rightarrow NNC^h) \rightarrow NC^h \rightarrow C^h$$

The motivator of this sound change is claimed in this paper to be phonetic: ease of articulation motivates the devoicing of the nasal portion of prenasalized stops, and the relatively low amplitude of the ensuing voiceless nasalization, which does not always rise above background noise, particularly in real-world situations, as opposed to ideal lab conditions, leads to speakers mishearing the sound as simply aspirated, which eventually leads them to produce the sound without nasalization.

4.3. Further Study

Considering the tendency for the 60 ms. VOT to lead to low rates of “prenasalized” identification when the recording also lacks nasalization, further study could focus on exactly how long the duration of the aspiration needs to be in order for Somali Chizigula speakers to identify a stop as

prenasalized by varying the VOT duration. A study which controls for aspiration length could then be done to see if this results in more consistent identification. Varying the nasal would be perhaps more difficult, but may lead to clearer information on how loud a nasal has to be in order to be heard. A study done including more participants might provide more conclusive results, or may show a tendency for different speakers to weight cues differently.

Another direction of interest would be a generational study, such as the Seoul Korean study done by Kang, which compares both the production of the nasalization and aspiration between older and younger speakers and the perception of the sounds between the same groups, to see if there is an apparent difference, indicating a transition in process. A large-scale corpus study would be unfeasible, due to the comparatively small population of available Somali-Chizigula speakers, but a smaller-scale study may yield interesting results if younger speakers appear to be weighting aspiration more heavily than older speakers.

REFERENCES

- Archangeli, D., Moll, L., & Ohno, K. (1998). Why not* NC. *CLS*, 34(1), 1-26.
- Boersma, Paul & Weenink, David (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4, retrieved March 25, 2015 from <http://www.praat.org/>
- Busa, M. G., & Ohala, J. J. (1995). Nasal loss before voiceless fricatives: a perceptually-based sound change. *Rivista di Linguistica*, 7, 125-144.
- Casali, R. F. (1995). NCs in Moghamo: prenasalized onsets or heterosyllabic clusters?. *Studies in African Linguistics*, 24, 151-166.
- Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34(3), 372-387.
- Cho, T., Jun, S. A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of phonetics*, 30(2), 193-228.
- Contini-Morava, Ellen. (1997). Swahili Phonology. *Phonologies of Asia and Africa: Including the Caucasus*. 841-860
- Dayley, Jon P., Mwaliko Mberwa, and Michal Temkin Martinez. (2016). *Chizigula of Somalia - English Dictionary*. Webonary.org. SIL International.
- Downing, L. J. (2005). On the ambiguous segmental status of nasals in homorganic NC sequences. *The internal organization of phonological segments*, 183-216.
- Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. *Origins of sound change: Approaches to phonologization*, 51-97.
- Han, M. S., & Weitzman, R. S. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/ and /ph, th, kh/. *Phonetica*, 22(2), 112-128.
- Hinnebusch, T. J. (1975). A reconstructed chronology of loss: Swahili class 9/10. *Proceedings of the Sixth Conference on African Linguistics*. 20, 32-41
- Huffman, M. K., & Hinnebusch, T. J. (1998). The phonetic nature of voiceless nasals in Pokomo: Implications for sound change. *Journal of African languages and linguistics*, 19(1), 1-19.
- Hyman, L.(2013). Enlarging the scope of phonologization *. *Origins of Sound Change: Approaches to Phonologization*. 3-28.
- Hyman, L. M. (1972). Nasals and nasalization in Kwa. *Studies in African linguistics*, 3(2), 167-205.
- Hyman, L. M. (2003). Segmental phonology. *The Bantu Languages*, 42-58.
- Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, 45, 76-90.

- Kim, M. R., Beddor, P. S., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, 30(1), 77-100.
- Kirby, J. P. (2010). *Cue selection and category restructuring in sound change* (Doctoral dissertation, The University of Chicago).
- Kurowski, K., & Blumstein, S. E. (1984). Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. *The Journal of the Acoustical Society of America*, 76(2), 383-390.
- Maddieson, I. (2003). The sounds of the Bantu languages. *The Bantu Languages*, 15-41.
- Maddieson, I. & Ladefoged, P. (1996). Nasals and Nasalized Consonants. *The Sounds of the World's Languages*. 102-136.
- Maddieson, I., & Ladefoged, P. (1993). Phonetics of partially nasal consonants. *Nasals, Nasalization and the Velum*, 5, 251-301.
- Mielke, J. (2003). The Diachronic Influences of Perception: Experimental Evidence from Turkish. *Annual Meeting of the Berkeley Linguistics Society* (Vol. 29, No. 1, pp. 557-567).
- Myers, S. (2002). Gaps in factorial typology: The case of voicing in consonant clusters. *Ms., University of Texas at Austin*, 1-35.
- Nurse, Derek and Philippon, Gerard. (2003). *The Bantu Languages*.
- Ohala, J. J. (1993). Sound change as nature's speech perception experiment. *Speech Communication*, 13(1-2), 155-161.
- Pater, J. (1999). Austronesian nasal substitution and other NC effects. *The prosody-morphology interface*, 310-343.
- Schadeberg, T. C. (2003). Historical linguistics. *The Bantu Languages*, 143-163.
- Tak, J. Y. (2003). Prenasalized consonants in Bantu. *음성음운형태론연구*, 9(2), 499-513.
- Tak, J. Y. (2011). Universals of Prenasalized Consonants. *Journal of Universal Language*, 12(2), 127-158.
- Takagi, Naoyuki, and Virginia Mann. (1995). The Limits of Extended Naturalistic Exposure on the Perceptual Mastery of English /r/ and /l/ by adult Japanese Learners of English. *Applied Psycholinguistics*, 16(4). 379-405.
- Temkin Martinez, Michal, & Boone, Haley. (2016). On the Presence of Voiceless Nasalization in Apparently Effaced Prenasalized Stops in Somali Chizigula. *The Journal of the Acoustical Society of America* 139.4, 2218-2218.
- Temkin Martinez, M, & Rosenbaum, V. (2017). Acoustic and Aerodynamic Data on Somali Chizigula Stops. *Africa's Endangered Languages: Documentary and Theoretical Approaches*, 427.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive science*, 34(3), 434-464.

Wang, Mengqian, & Boone, Haley. (2017). Acoustic Evidence for Voiceless Prenasalization in Somali Chizigula Stops. *Paper presented at the 2017 SECOL conference*.